Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 35 (2014) 281 – 289

Procedia Computer Science

18th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems - KES2014

Soft approach to identification of cohesive clusters in two gene representations

Michał Kozielski*, Aleksandra Gruca

Silesian University of Technology, Institute of Electronics, Akademicka 16, 44-100 Gliwice, Poland

Abstract

The approach to identify clusters of genes represented both by expression values and Gene Ontology annotations, where cluster membership should not be in conflict with any of the representations is presented in the paper. The method enables to identify the genes that are differently clustered in different representations, what can lead to further analysis and interesting conclusions. The approach is based on the fuzzy clustering algorithms and the notion of proximity as the aggregation operation at the higher level than similarity matrices is performed. The approach is verified on two datasets: a small synthetic and real-world gene dataset.

Peer-review under responsibility of KES International.

Keywords: proximity; fuzzy clustering; fuzzy aggregation; gene expression; Gene Ontology;

1. Introduction

Gene Ontology (GO)[1] is a widely used knowledge base that is continuously developed and corrected according to the newest knowledge derived from different biological experiments. Therefore, Gene Ontology is an important source of knowledge that can be utilised in further analysis.

GO enables annotation of gene products to ontology terms representing biological process, molecular function or biological component. It is modelled as a directed acyclic graph where ontology terms are graph nodes and the edges are defined by the relations between terms.

One of important type of analysis in the field of bioinformatics is looking for gene groups characterised by similar expression patterns identified in a micro-array experiment. It is possible to enrich the information represented by a single micro-array experiment by the information collected throughout the years in Gene Ontology. Thus, it can be valuable to combine these two gene representations: gene expression and Gene Ontology, along the clustering process.

The goal of this work is to present an approach to identify clusters of genes represented both by expression values and GO annotations, where cluster membership should not be in conflict with any of the representations. In this way it

* Corresponding author.
E-mail address: michal.kozielski@polsl.pl

Peer-review under responsibility of KES International.
doi:10.1016/j.procs.2014.08.108

is also possible to identify the genes that are differently clustered in different representations, what can lead to further analysis and interesting conclusions.

Such combination of the two representations of the same data objects can be performed at the level of:

- similarity (or dissimilarity) matrices applying aggregation operation,
- clustering process, where a clustering algorithm can manage multi-represented data,
- partitions resulting from a clustering process, where ensemble methods can be applied in order to calculate a resulting, single partition.

We focus on the level of similarity matrix and clustering algorithm in this work. Therefore, it is important to verify these approaches out of the mentioned above. The Proximity-based Fuzzy C-Means (PFCM) algorithm was introduced by Pedrycz et al.[2] to identify the global structure of the data objects described by two representations. The method was further extended to reconcile a chosen number of representations[3] and it was also applied to gene data analysis[4].

The global density-based clusters of the data having many representations can be identified by means of the method presented in[5]. Operations of *intersection* and *union* were utilised in this work in order to identify data objects creating dense clusters in each or any (respectively) data representations.

The clustering approaches mentioned above are not suitable to our task. The PFCM algorithm is looking for a consensus partition covering all the data objects, whereas, we want to identify objects that are members of significantly different clusters in different representations. Density-based approach is looking for the clusters which are dense in each representation, whereas in our case it is enough if the object-cluster membership is not divergent.

Also aggregation of the similarity matrices could be an interesting solution for the given task. The weighted combination of the two gene representations was presented in[6]. Such approach enables to point which representation should have a stronger impact on the results, what is, however, not so obvious. Another approach presented in[7] introduced fuzzy aggregation of gene expression and Gene Ontology based similarity matrices. The similarity matrices aggregation will be referred further as a reference result of the experiments.

The approach that is a contribution of this paper requires combination of the two representations what is performed by means of the aggregation operation but at the higher conceptual level than similarity matrices. Thus, the notion of proximity and fuzzy clustering algorithms were applied in this task.

The structure of this work is as follows. Section 2 presents the approach that is introduced in this work. The analysis of two datasets (synthetic and real-world) and discussion of the results is presented in section 3. Summary and final conclusions are presented in section 4.

2. Proposed approach

The main blocks of the approach introduced in this work are presented in the fig. 1. This method can be seen as a hybrid approach in the context of the points listed in the previous section, as it combines the features of an aggregation of similarity matrices and ensemble methods applied to the resulting partitions.

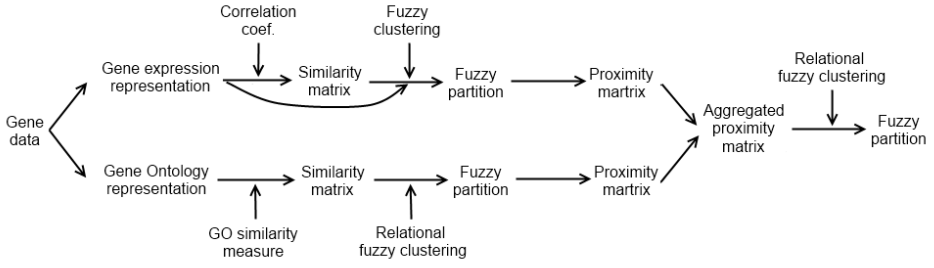


Fig. 1: The approach diagram

The data that are analysed in this work are complex what influences the way their similarity is calculated and the choice of the clustering method to be applied. The first issue of the approach is calculation of similarity/distance matrix of the analysed data in each representation - gene expression and Gene Ontology. The similarity in gene expression representation is typically calculated as a Pearson correlation coefficient[8], as gene expression values are time series of real numbers.

Genes represented by annotations to the terms of Gene Ontology graph are usually modelled by annotation table. This is a binary matrix where genes and ontology terms are the two dimensions and an annotation is represented as 1 value. The similarity in Gene Ontology representation can be calculated by means of one out of many existing measures[9]. A number of the methods is based on the two steps: at first Gene Ontology term similarity is calculated and next the similarity of genes annotated to these terms is calculated. Among term similarity measures we can mention e.g. semantic methods[9] which are based on a common ancestor identification (e.g. Resnik method[10]) or path-based method[11] which is applied directly on the GO graph. The previous comparative studies[12] show that the path-based measure gives the results of a good quality. This measure is based on a summed length $\gamma(a_i, a_j)$ of the shortest paths from a pair of GO terms a_i and a_j to their lowest common ancestor and is calculated according to the following formula:

$$s(a_i, a_j) = e^{-f\,\gamma(a_i,a_j)}, \tag{1}$$

where f is a constant.

Among different term-based gene similarity measures[13,14] the measure called further Avg-max, which calculates average of maximal term similarity gives good results when applied together with path-based term similarity measure[15]. The Avg-max measure is calculated according to the following formula:

$$s_G(g_k, g_p) = (m_k + m_p)^{-1}\Big(\sum_i \max_j(s(a_i, a_j)) + \sum_j \max_i(s(a_i, a_j))\Big), \tag{2}$$

where m_k and m_p are the number of annotations of genes g_k and g_p respectively, a_i and a_j belong to the term sets describing genes g_k and g_p respectively.

When the gene similarity matrices are calculated it is possible to apply to them a fuzzy aggregation as it was presented in[7]. However, we will apply clustering algorithm at first, in order to identify clusters of genes that are similar to each other.

The main algorithm that is utilised in the approach is fuzzy clustering algorithm as fuzzy partition matrix is required to perform further steps of the analysis. The most popular fuzzy clustering algorithm is Fuzzy C-Means (FCM). This is partitional, optimisation algorithm that requires a number of groups to be defined as a parameter. This method iteratively calculates a distance between the data and a calculated (as a mean value) cluster prototype. It is possible to modify FCM so that it will use Pearson correlation as a basis of distance function and it will be applicable to gene expression data. However, such approach is not applicable to Gene Ontology data where relational approach is required. Additionally, in this case the cluster prototypes cannot be artificial objects calculated as a mean value but they have to be selected from the analysed set of genes. Thus, relational method named Robust Fuzzy C-Medoids (RFCMdd) introduced by Krishnapuram[16] is suitable for this task.

As a result of fuzzy clustering we receive two fuzzy partitions, where each of them is defined as $\mathbf{U} = [u_{ik}]$, $i = 1, 2, .., c$, $k = 1, 2, .., N$, where N is a number of genes, c is a number of clusters and u_{ik} is a membership value of k-th data object to i-th cluster. Each of the partitions is calculated on the basis of the same data but of different representation.

It would be possible to apply ensemble methods[17,18] to calculate a single resulting partition in the next step. However, we would like to perform additional analysis at the higher, granular level. Therefore, in order to receive a uniform representation of the data, which is independent on the number of clusters identified in a clustering process, a proximity matrix is calculated for each representation. Proximity matrix $\mathbf{P} = [p_{kl}]$ is calculated on the basis of fuzzy partition matrix $\mathbf{U}$ according to the following formula[2]:

$$p_{kl} = \sum_{i=1}^{c} \min(u_{ik}, u_{il}) \tag{3}$$

The notion of proximity was utilised in the clustering algorithms before, e.g., in Proximity-based Fuzzy C-Means (PFCM)[2]. In a given work we are, however, not interested in reconciling the partitionings as they are. The goal

of the work is to identify clusters of genes, where cluster membership should not be in conflict with any of the representations. Possibly the genes presenting divergent cluster membership across representations should be also identified for further analysis. Therefore, it was decided to perform a proximity matrices aggregation at first and then to apply clustering algorithm on the basis of a common proximity matrix.

The resulting proximity matrix is calculated as $\mathbf{P}_A = A(\mathbf{P}_1, \mathbf{P}_2)$, where A is an aggregation function. The *min* function was chosen in the presented approach, as this operator is beneficial because the genes belonging to different clusters can be identified as having a low resulting proximity values. Therefore, in order to identify such genes it is needed to calculate a vector $\hat{\mathbf{p}}_A = [\hat{p}_l]$ such that:

$$\hat{p}_l = \max_{\forall k=1..N, k\neq l} p_{kl} \tag{4}$$

for $l = 1, ..., N$. The low $\hat{p}_l$ value indicates that a given l-th gene belongs to different clusters in different representations.

The final step of the approach consists of the application of the relational clustering method to the resulting proximity matrix in order to identify a final partition taking both representations into consideration. If there are any genes presenting divergent cluster membership across representations identified it is possible to investigate what is the reason of such behaviour. The identified discrepancies can be caused, e.g., by incorrect data values that were not identified during data cleaning process or by the new dependencies that were previously uncovered.

3. Experimental results

The experiments that were performed were conducted on two datasets - synthetic example dataset and real-world gene dataset. The experiments enable evaluation of the approach presented in section 2. The results of the analysis were compared with the approach based on similarity matrices aggregation by means of *min* function, that was utilised in [7].

3.1. Synthetic dataset

Synthetic example dataset consists of 9 data objects described in two representations based on two dimensional (Euclidean) space. The dataset was designed in order to present clearly how the consecutive steps of the approach are calculated and what are the results of the method. The two representations of the synthetic dataset are presented in fig. 2. The data objects create two easily visually identifiable clusters of different densities. One of the objects, having id 9 (filled in with a solid colour), changes its cluster membership between the representations. This characteristic of object 9 should be identified by the introduced approach.

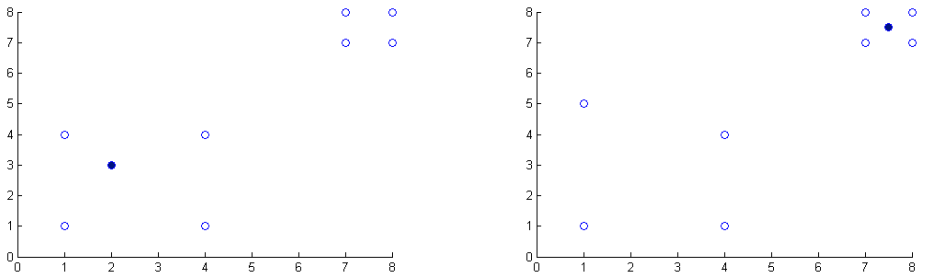


Fig. 2: The two representations of example dataset

FCM clustering algorithm was applied as the first step of the method to identify a pair of clusters in each representation (a and b). The number of clusters was set to $c = 2$, the fuzzification coefficient was set to $m = 2$. The resulting fuzzy partition matrices are presented in table 1. The object membership values presented in table 1 show that the

Table 1: Fuzzy partition matrices ($c = 2$) calculated for example dataset

Object id	1	2	3	4	5	6	7	8	9
Representation a	0.953	0.934	0.909	0.826	0.010	0.009	0.010	0.009	0.994
	0.047	0.066	0.091	0.174	0.990	0.991	0.990	0.991	0.006
Representation b	0.053	0.143	0.084	0.153	0.990	0.990	0.990	0.990	1.000
	0.947	0.857	0.916	0.847	0.010	0.010	0.010	0.010	0.000

clustering algorithm identified the clusters existing in each representation very clearly.

In the next step, the proximity matrices were calculated on the basis of the partition matrices and then, the aggregation was applied to combine the proximity values into a resulting matrix $\mathbf{P}_{min}$. The resulting proximity matrix is presented in table 2.

Additionally, in table 3 an aggregated similarity matrix (min function was applied) is presented in order to compare what are the results when aggregation is performed at the level of similarity matrices and at the level of proximity matrices.

Table 2: Proximity matrix aggregating the two representations

Object id	1	2	3	4	5	6	7	8	9
1	1.000	0.911	0.956	0.873	0.056	0.056	0.056	0.056	0.053
2	0.911	1.000	0.941	0.892	0.075	0.075	0.075	0.075	0.143
3	0.956	0.941	1.000	0.917	0.094	0.094	0.094	0.094	0.084
4	0.873	0.892	0.917	1.000	0.163	0.164	0.164	0.163	0.153
5	0.056	0.075	0.094	0.163	1.000	1.000	1.000	1.000	0.015
6	0.056	0.075	0.094	0.164	1.000	1.000	1.000	1.000	0.015
7	0.056	0.075	0.094	0.164	1.000	1.000	1.000	1.000	0.015
8	0.056	0.075	0.094	0.163	1.000	1.000	1.000	1.000	0.015
9	0.053	0.143	0.084	0.153	0.015	0.015	0.015	0.015	1.000

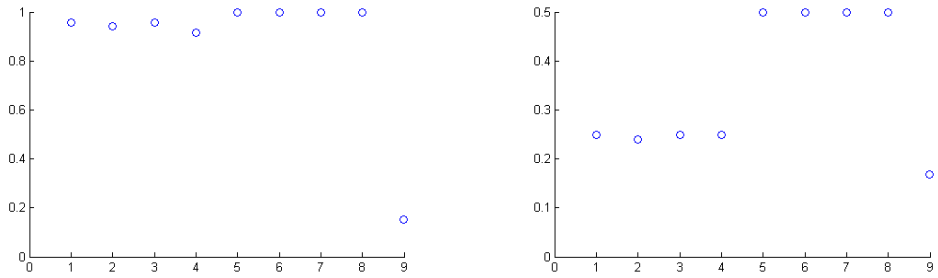
Table 3: Similarity matrix aggregating the two representations

Object id	1	2	3	4	5	6	7	8	9
1	1.000	0.200	0.250	0.191	0.098	0.098	0.105	0.092	0.098
2	0.200	1.000	0.167	0.240	0.122	0.116	0.130	0.110	0.126
3	0.250	0.167	1.000	0.250	0.116	0.122	0.130	0.110	0.119
4	0.191	0.240	0.250	1.000	0.167	0.167	0.191	0.150	0.168
5	0.098	0.122	0.116	0.167	1.000	0.414	0.500	0.500	0.124
6	0.098	0.116	0.122	0.167	0.414	1.000	0.500	0.500	0.122
7	0.105	0.130	0.130	0.191	0.500	0.500	1.000	0.414	0.135
8	0.092	0.110	0.110	0.150	0.500	0.500	0.414	1.000	0.114
9	0.098	0.126	0.119	0.168	0.124	0.122	0.135	0.114	1.000

The maximal proximity vector $\hat{\mathbf{p}}_{min}$ and analogous maximal similarity vector which are calculated on the basis of matrices from tables 2 and 3 respectively are visualised in fig. 3 (a) and (b).

The values of the maximal proximity vector visualised in fig. 3 (a) enable to identify the data object of radically different membership between the representations very easily. The task is not so obvious in case of the aggregated similarity matrices. As it can be seen in fig. 3 (b), that except object 9, there are four data objects having also low value of maximal similarity. Therefore, it is not clear which objects belong to different clusters in different representations and which objects just belong to a cluster of a low density.

When object 9 is identified as having strongly divergent cluster membership for different representations it is possible to investigate what is the reason of such behaviour. Next, knowing the specificity of the dataset, it can be clustered in order to receive a single partition based on the two representations.



(a) (b)

Fig. 3: The aggregated proximity (a) and similarity (b) values calculated for the objects of synthetic dataset

3.2. Real-world gene dataset

The second dataset that is used in the presented approach evaluation consists of budding yeast *S. cerevisiae* genes, thus it will be further referred as Yeast dataset. The expression profiles of these genes were measured during several different DNA microarray experiments and their analysis was presented in [8]. The dataset consists of 274 genes, each described by 79 expression values. On the basis of this gene expression representation 10 groups of genes were identified in [8].

The same set of genes can be described by means of the annotations to Gene Ontology terms. We focused on *Biological process* subontology in the analysis presented below. Therefore, the dataset in GO representation consists of 274 genes described in the space of 248 GO terms. As it was described in the previous sections GO representation has the form of the binary annotation table.

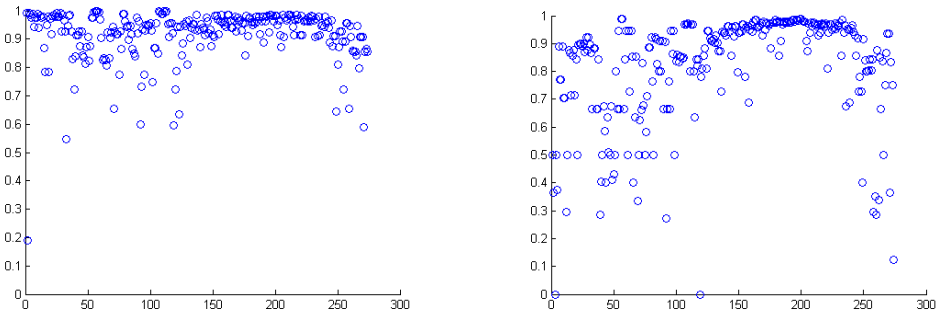
In order to analyse Yeast dataset it was needed to apply different methods than in case of synthetic dataset. Gene expression representation was analysed by means of the FCM algorithm where Pearson correlation coefficient is applied to calculate the distance between the genes and cluster prototypes. Eisen et al. identified in their work [8] 10 clusters on the basis of gene expression representation of Yeast dataset. Therefore, it was decided to look for the best partition for the number of clusters set to $c \in \{7, 8, 9, 10, 11, 12, 13\}$ and the fuzziness coefficient set to $m = 2$. The quality index presented in [19] was applied to asses which fuzzy partition is of the best quality. The partition identified for $c = 13$ was chosen as the best one. Additionally the partition for $c = 10$ was also calculated in order to compare the results. Due to very little differences in proximity values calculated for these two partitions, the results for $c = 10$ are further presented.

The similarity of genes in Gene Ontology representation was calculated by means of path-based term similarity method (eq. 1) and Avg-max term-based gene similarity (eq. 2). The Robust Fuzzy C-Medoids (RFCMdd) algorithm was applied in order to identify fuzzy partition of genes in this representation. Again, the best partition was searched when the number of clusters set to $c \in \{7, 8, 9, 10, 11, 12, 13\}$ and the fuzziness coefficient was set to $m = 1.1$. The best partition was identified for $c = 12$ clusters.

On the basis of the partition matrices the proximity matrices were calculated and the aggregation was applied to combine the proximity values into a resulting proximity matrix. Again an aggregated similarity matrix was calculated in order to compare what are the results when aggregation is performed at the level of similarity matrices and at the level of proximity matrices.

The maximal proximity vector and analogous maximal similarity vector were calculated and visualised in fig. 4 (a) and (b) respectively, in order to identify genes that belong to different clusters in different representations. The plots presented in fig. 4 show that in case of aggregated similarity matrix (fig. 4 (b)) there are much more genes that could be suspected to belong to different clusters in different representations. This is similar conclusion to the one drawn on the basis of synthetic dataset. Fig. 3 (b) showed that more data objects were suspected but only one belonged to divergent clusters for different representations.

In order to verify if the objects identified in fig. 4 (a) as having low value of maximal proximity are rightly suspected, the biological analysis was performed. Two genes having the lowest value of maximal proximity were



(a) (b)

Fig. 4: The aggregated proximity (a) and similarity (b) matrix of Yeast dataset

investigated and compared with the members of the clusters identified in the expression representation. The three genes characterised by the highest correlation with the investigated one were chosen as the representative members of the cluster.

Gene EXO1 has id 2 in the Yeast dataset and it stands out in the fig. 4 (a) as its maximal proximity value equals $\hat{p}_2 = 0.19$. The expression of this gene is most correlated with the expression of genes: BNR1 (id 3), SPC42 (id 5), CNM67 (id 4) and the values of correlation are presented in table 4. All these genes belong to cluster 1 identified by Eisen et al.[8]. Although the expression of the genes presented in table 4 is highly correlated, it can be found that the

Table 4: Correlation of the investigated genes

	BNR1 (id 3)	SPC42 (id 5)	CNM67 (id 4)
EXO1 (id 2)	0.944	0.917	0.912

gene 2 and the genes 3, 5 and 4 play role in different biological processes. In contrast to EXO1 (id 2) the other three genes (BNR1, SPC42, CNM67) are associated with the processes connected to spindle pole body.

The second gene that undergone a more detailed analysis is RPN10, id 33 in Yeast dataset. This gene has the second lowest value of maximal proximity (see fig. 4 (a)), which is equal $\hat{p}_{33} = 0.547$. The expression of this gene is most correlated with the expression of genes: PRE2 (id 28), PRE3 (id 27), PUP1 (id 30) and the values of correlation are presented in table 5. All these genes belong to cluster 2 identified by Eisen et al.[8]. Again, although the expression

Table 5: Correlation of the investigated genes

	PRE2 (id 28)	PRE3 (id 27)	PUP1 (id 30)
RPN10 (id 33)	0.812	0.810	0.804

of the genes presented in table 5 is highly correlated, we can show differences between gene 33 and the genes 28, 27 and 30. Gene RPN10 (id 33) is a subunit of the 19S RP of the 26S proteasome, whereas the other three genes (PRE2, PRE3, PUP1) are the subunits of the 20S proteasome.

Knowing and understanding the characteristics of the data that are analysed it is possible to apply final clustering method in order to calculate a single partition of the double-represented genes. The investigated genes could be rejected or processed in a special way, however it was decided to present the analysis of the whole dataset as an example. Additionally, instead of identifying the final partition by means of clustering algorithm the dissimilarity matrices visualisation was applied. Fig. 5 contains visualisations of the proximity and similarity matrices that were calculated for Yeast dataset. They enable a comparison of both proximity and similarity approaches, as well as each representation and their aggregation.

Fig. 5 shows that in case of proximity matrices the resulting aggregation is impacted by both gene expression and GO representations. In case of similarity matrices the resulting aggregation is dominated by GO representation (both visualisations are indistinguishable).

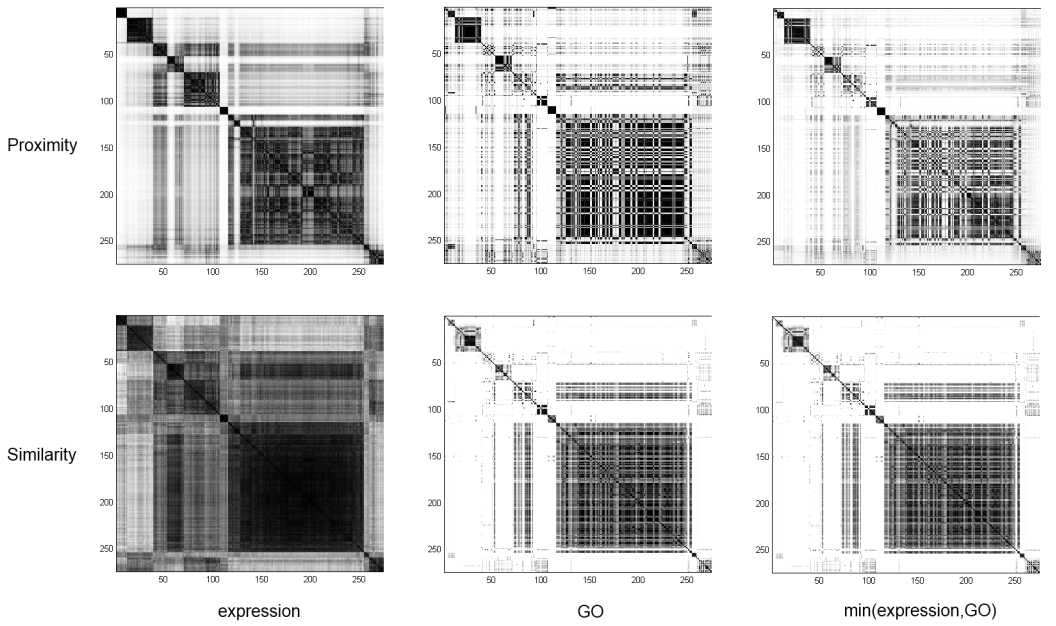


Fig. 5: Dissimilarity visualisations calculated on the bases of similarity and proximity matrices for expression, Gene Ontology and aggregated representation of Yeast dataset

In this way another advantage of proximity analysis is visible. Although the similarity matrices of both gene representations are normalised to [0,1] range one of the representations can dominate the aggregated (by min function) representation due to the lower similarity values. Proximity analysis is based on a higher conceptual level and, therefore, this issue does not occur for this approach.

4. Conclusions

The work presents a new approach to identify clusters of genes represented both by expression values and GO annotations. The clusters are built according to the assumption that the cluster membership should not be in conflict with any of the representations. The approach is based on the fuzzy clustering algorithms and the notion of proximity as the aggregation operation at the higher level than similarity matrices is performed.

The method characteristics were explained on a synthetic dataset and verified and evaluated on real-world gene dataset. The comparison with the method where the representations are aggregated at the level of similarity matrices was also performed.

The method enables to identify the genes that belong to divergent clusters in different representations, what can lead to further analysis and interesting conclusions. Two such example genes were identified and their divergent characteristics in the two representations were described.

It was also shown that the aggregation of the representations at the higher conceptual level by means of the application of proximity has additional advantage. It has better ability to incorporate the features of both representations than similarity analysis where one representation can dominate the aggregated representation.

Acknowledgments

The work was supported by National Science Centre (decision DEC-2011/01/D/ST6/07007). The work was performed using the infrastructure supported by POIG.02.03.01-24-099/13 grant: GeCONiI-Upper Silesian Center for Computational Science and Engineering.

References

1. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., et al. Gene ontology: tool for the unification of biology. *Nature genetics* 2000;**25**(1):25–29.
2. Pedrycz, W., Loia, V., Senatore, S.. P-fcm: a proximityï£¡based fuzzy clustering. *Fuzzy Sets and Systems* 2004;**148**(1):21–41.
3. Pedrycz, W.. Proximity-based clustering: A search for structural consistency in data with semantic blocks of features. *Fuzzy Systems, IEEE Transactions on* 2013;**21**(5):978–982.
4. Gruca, A., Kozielski, M., Sikora, M.. Fuzzy clustering and gene ontology based decision rules for identification and description of gene groups. In: *Man-Machine Interactions*. Springer; 2009, p. 141–149.
5. Kailing, K., Kriegel, H.P., Pryakhin, A., Schubert, M.. Clustering multi-represented objects with noise. In: *Advances in Knowledge Discovery and Data Mining*. Springer; 2004, p. 394–403.
6. Kustra, R., Zagdanski, A.. Incorporating gene ontology in clustering gene expression data. In: *Computer-Based Medical Systems, 2006. CBMS 2006. 19th IEEE International Symposium on*. IEEE; 2006, p. 555–563.
7. Havens, T.C., Keller, J.M., Popescu, M., Bezdek, J., MacNeal Rehrig, E., Appel, H., et al. Fuzzy cluster analysis of bioinformatics data composed of microarray expression data and gene ontology annotations. In: *Fuzzy Information Processing Society, 2008. NAFIPS 2008. Annual Meeting of the North American*. IEEE; 2008, p. 1–6.
8. Eisen, M., Spellman, P., Brown, P., Botstein, D.. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 1998;**95**(25):14863–14868.
9. Pesquita, C., Faria, D., Falco, A., Lord, P., Couto, F.. Semantic similarity in biomedical ontologies. *PLoS Comput Biol* 2009;**5**(7):1–12.
10. Resnick, P.. Semantic similarity in a taxonomy: An information-based measure and its application to problems of ambiguity in natural language. *J Artif Intell Res* 1999;**11**:95–130.
11. Al-Mubaid, H., Nagar, A.. Comparison of four similarity measures based on go annotations for gene clustering. In: *IEEE Symposium on Computers and Communications, ISCC 2008*. 2008, p. 531–536.
12. Gruca, A., Kozielski, M.. Correlation of genes similarity measures based on go terms similarity and gene expression values. In: *Man-Machine Interactions 2*. Springer; 2011, p. 137–144.
13. Azuaje, F., Wang, H., Bodenreider, O.. Ontology-driven similarity approaches to supporting gene functional assessment. In: *Proc. of the ISMB'2005 SIG meeting on Bio-ontologies*. Mitchigen, USA; 2005, p. 9–10.
14. Wang, H., Azuaje, F., Bodenreider, O., Dopazo, J.. Gene expression correlation and gene ontology-based similarity: an assessment of quantitative relationships. In: *Proc. IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology CIBCB '04*. 2004, p. 25–31.
15. Kozielski, M., Gruca, A.. Visual comparison of clustering gene ontology with different similarity measures. *Studia Informatica* 2011; **32**(2A):169–180.
16. Krishnapuram, R., Joshi, A., Nasraoui, O., Yi, L.. Low-complexity fuzzy relational clustering algorithms for web mining. *Fuzzy Systems, IEEE Transactions on* 2001;**9**(4):595–607.
17. Strehl, A., Ghosh, J.. Cluster ensembles—a knowledge reuse framework for combining multiple partitions. *The Journal of Machine Learning Research* 2003;**3**:583–617.
18. Topchy, A., Jain, A.K., Punch, W.. Clustering ensembles: Models of consensus and weak partitions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 2005;**27**(12):1866–1881.
19. Łęski, J., Czogała, E.. A new artificial neural network based fuzzy inference system with moving consequents in if–then rules and selected applications. *Fuzzy Sets and Systems* 1999;**108**(3):289–297.